

Rationale for the MR Sign Concordance Test

The ideal scenario for a Mendelian Randomization (MR) analysis is in which there are multiple independent SNPs instrumenting the exposure. Here, each SNP acts as a proxy for a randomised controlled trial, and each SNP can provide an independent causal effect estimate through the Wald ratio. The effect estimate will be unbiased if, for example, the SNP only influences the outcome through the exposure (i.e. there is no horizontal pleiotropy). If all SNPs provide unbiased causal effect estimates, or if bias across all SNPs is random in direction and magnitude, meta-analysing across the effect estimates will give an overall unbiased estimate of the causal effect that is more precise than any individual SNP.

Proving that the bias from each SNP is absent or perfectly balanced is not possible, and if this untestable assumption is not satisfied then the average effect across all SNPs will be inevitably biased. Heterogeneity tests can determine whether there are differences between each SNP's estimate that goes beyond that expected due to their standard errors. When substantial heterogeneity is present, it is prudent to follow up with additional meta-analysis strategies that are robust to different patterns of pleiotropy manifesting across the SNPs. However, generally these strategies operate by averaging across all SNPs (e.g. IVW, MR-Egger), or averaging across a subset of SNPs (e.g. weighted median, weighted mode).

In our analysis we performed MR of 116 traits against 345,109 DNA methylation sites (42,103,298 analyses). Systematically searching for putative results of interest requires prioritising associations that are not due to obvious violations. A chief source of bias arises if one of many instruments for a trait happen to be close to a variant that has a very large effect on the DNA methylation site. This can arise under a range of common scenarios, for example

- Being close to a *cis* mQTL for the DNA methylation site
- Being located within the *MHC* region, which has large effects on many sites
- Being located close to a SNP influencing smoking, which has consequently large influence on DNA methylation sites.

The available methods, which average across the Wald ratios from all instruments, are all susceptible to SNPs that might have very large effects on the outcome. This is even true for the weighted median and weighted mode estimators, because the weights are based on the precision of the SNP-outcome association, and a single SNP could comprise half or the majority of the weight if it has a particularly large effect.

Standard methods attempt to estimate the causal effect of the exposure on the outcome, and then report the strength of evidence by estimating the precision of the estimate and often evaluating if the estimate is substantially different from 0. Here, the qualitative evaluation of whether there is a causal effect (and its direction) manifests from the estimate of the effect size.

Because the effect estimate, a form of an average across a set of SNPs, is liable to be skewed, we developed a method that evaluates the qualitative question of whether there is a

causal effect (and its direction) independent of the estimated effect size. This was simply achieved by inferring the extent to which the causal directions of the Wald ratios from all SNPs are consistent. If the trait truly causes the DNA methylation site then in general the effect estimates from each of the SNPs should have the same sign, and each SNP's contribution to this concordance is equal. Under the null hypothesis of no causal effect, we expect 50% of the signs of the Wald ratios to be positive and 50% to be negative. The MR sign concordance test (MR-SCT) simply performs a binomial test to see if the proportion of positive signs is more (in the case of a positive effect) or less (in the case of a negative effect) than expected by chance.

The MR-SCT doesn't provide a causal effect estimate. It returns the proportion of Wald ratios that are positive. An MR-SCT estimate of 1 would indicate very strong agreement for a positive effect of the exposure on the outcome, and a value of 0 would indicate very strong agreement for a negative effect. The corresponding p-value can be used for evaluating the extent to which the proportion is different to what we expect by chance.